

GMDR User Manual

Version 1.0

Oct 30, 2011

GMDR is a free, open-source interaction analysis tool, aimed to perform gene-gene interaction with generalized multifactor dimensionality methods. **GMDR** is being developed by Guo-Bo Chen.

Contact Information

Guo-Bo Chen

Email: guobo.chen@uq.edu.au

Address:

Queensland Brain Institute

The University of Queensland

St Lucia, QLD 4072, Australia

Content

1.System Requirements	5
Begins GMDR with java –jar gmdr.jar	5
Use Xmx for large dataset	5
2.Options of GMDR	5
3.Help & Version	5
4.Data Format	6
Text PED files (ped, map)	6
Binary PED files (bed, bim, fam)	6
5.Detect Interactions.....	7
--cc (unrelated case-control sample)	7
--pi (pedigree-based method I by generating pseudo individuals).....	7
--pii (pedigree-based method II by within family controls)	7
--ui (unified method I with family and unrelated samples, exchangeable founders).....	8
--uii (unified method II with family and unrelated samples, unexchangeable founders)	8
--x (detect interaction between candidate regions)	8
6.Evaluating Empirical p values with permutation	8
7.SNP selection	9
--snp (include or exclude SNPs).....	9
--snpf (include or exclude SNPs through a file).....	9
--chr (include or exclude chromosomes)	10
--window (include SNPs through windows)	10
--bg (keep certain SNP patten when detecting interactions).....	10
--maf (threshold for excluding SNPs)	10
8.Individual selection	11
--exfam (exclude families)	11
--exfamf (exclude families through a file)	11
--exnosex (exclude individuals whose sexes were unknown).....	11
--male (include males only).....	11
--female (include females only)	11
--geno (threshold for genotype missing rates)	11
9.Phenotypes and covariates	11
--phenotype (specify a phenotype file)	11
--reg (method for building score statistics).....	12
10.Cross-validation options.....	12
--cv (fold for cross-validation)	12
--seed	12
--tie (classification for a tie group).....	13

11. Output options.....	13
--out (the root name for the output files)	13
--verbose (complete the output information)	13
--train (threshold for training accuracy for output)	13
--test (threshold for testing accuracy for output)	14
--p (threshold for p value for output)	14
--vc (threshold for variance contributed for output)	14
12. Estimate the time required to complete the analysis.....	14
13. Reduce computational burden	14
--thin.....	14
14. GMDR and HPC	15
--slice	15
15. Miscellanea	16
--ss	16
--missingallele	16
--missingpheno.....	16
16. References.....	17
17. List of options.....	18

GMDR is a free, open-source interaction analysis tool, aimed to perform gene-gene interaction with generalized multifactor dimensionality methods. **GMDR** is being developed by Guo-Bo Chen.

1. System Requirements

Begins GMDR with java -jar gmdr.jar

GMDR runs in operating systems, such as Windows, Mac OS, Linux/Unix, in which a Java Virtual Machine is preinstalled. GMDR command often starts with

```
java -jar gmdr.jar --bfile example
```

Use Xmx for large dataset

By default, java set For the analysis of a big dataset (over 100M)

```
java -jar -Xmx1g gmdr.jar --bfile example
```

-Xmx is entailed with the memory size want to claim. Xmx1g claims 1 gigabyte space. It is often necessary to specify -Xmx when running a large dataset.

Notes: it should be noticed that there is only one dash "-" preceding both "jar" and "Xmx" because they are built-in options for java. For the GMDR options, always two dashes "--" precede them.

2. Options of GMDR

GMDR has two kinds of options: one kind of them takes one or more than one parameter, whereas the other not. For the first kind, such as "--bfile example" and "--perm 100" having parameters "example" and "100", respectively, the user should set parameters accordingly ; some options can have more than one parameter, such as "--snp snp0 snp1", "--chr 1 2". The other kind of options works like switches. They either turn on or turn off of certain function of GMDR. For example, "--verbose", "--ss", "--cc" are all switches and do not take any parameters.

This manual uses color theme to distinguish these two kinds of options. We use gray color highlight switch options.

3. Help & Version

"--help" and "--version" are two very basic options. "--help" option gives a brief view of the options built inside GMDR, and probably the first option should be tried when the user start to use GMDR

```
java -jar -Xmx1g gmdr.jar --help
```

The other thing the user is better to try is --version. When a much newer version is released, the user can download a new one

```
java -jar -Xmx1g gmdr.jar --version
```

4. Data Format

GMDR supports datasets organized in the pedigree format (PED). And the data can be either in text format (X.ped file and X.map files) or binary format (X.bed, X.bim, and X.fam files).

Text PED files (ped, map)

A PED file is white-space (space or tab) delimited: the first six columns are mandatory:

Family ID

Individual ID

Paternal ID

Maternal ID

Sex (1=male; 2=female; other=unknown)

Affection (1=unaffected; 2=affected; 0=unknown)

The IDs are alphanumeric: the combination of one's family and individual ID should uniquely identify an individual. The affection status is numerical and coded 1 for unaffected and 2 for affected (do not put any phenotype other than affection status at this column). Each SNP marker is represented in a pair of alleles. GMDR supports biallelic markers only. Missing alleles are coded as "0" by default, but it can be customized by specifying option "--missingallele".

By default, each line of the MAP file describes a single marker and must contain exactly 4 columns:

chromosome (1-22, X, Y or 0 if unknown)

SNP identifier, such as rs number.

Genetic distance (morgans)

Base-pair position (bp units)

Base-pair positions are expected to correspond to positive integers within the range of typical human chromosome sizes, but for basic interaction testing, the genetic distance column and base-pair position column can be set at 0 or arbitrary integer. Although SNP names can be any characters except white-space (spaces or tabs), but it is suggested to use alphanumeric.

GMDR reads in a ped file and a map file in two ways, depending on whether these two files have the same root name. When they do

```
java -jar gmdr.jar --file example
```

in this way, GMDR expects two files, example.ped and example.map, the same root name "example" shared. Once the root names differs, a ped file and a map file should be specified respectively

```
java -jar gmdr.jar --ped example.ped --map mymap.map
```

or

```
java -jar gmdr.jar --file example --map mymap.map
```

Binary PED files (bed, bim, fam)

To accommodate large datasets, such as GWAS data, GMDR suggests binary PED files. Binary PED files currently are often provided for those such as stored in dbGap. This format will store the pedigree information in *.fam file and an extended MAP file, *.bim, which contains information about the allele names. For how to create these files from text PED files, please refer to PLINK.

As with text PED files, when the root name of the binary PED files are the same, GMDR reads them with “--bfile” option:

```
java -jar gmdr.jar --bfile example
```

Once they do not share a common root name, the users should specify each of them respectively:

```
java -jar gmdr.jar --bim example.bim --bed example1.bed --fam example2.fam
```

For details of binary PED files, the users can refer to the web page of PLINK.

5. Detect Interactions

GMDR is enriched in options for detecting gene-gene interaction in different design, such as case-control design, pedigree-based design, and unified design pooling unrelated and family samples together. The examples invoked below assume the availability of example.bed, example.bim and example.fam.

--cc (unrelated case-control sample)

For case-control design

```
java -jar gmdr.jar --bfile example
```

which detects digenic interaction for all SNPs involved. This command performs exactly the same as

```
java -jar gmdr.jar --cc --bfile example
```

By default, GMDR detects interactions for unrelated individuals. For the theory underlying this option, please refer to *Am J Hum Genet 80:1125-37* [Lou, et al. 2007].

Mark: for a PED file containing nuclear families and case-control samples, this --cc option will pool founders and case-control sample together and calculate interaction. **Once --order option is not used, the default value is of 1.** To detect 2-locus interaction, specify option “--order 2”:

```
java -jar gmdr.jar --cc --bfile example --order 2
```

It detects digenic interactions among all SNPs in example.bim. **By default, it is 5-fold cross-validation.**

--pi (pedigree-based method I by generating pseudo individuals)

For pedigree-based design, **GMDR currently can handle nuclear families only.** Option --pi implements pedigree method I, which constructs within family pseudo individual as genetic control to each sibling.

```
java -jar gmdr.jar --pi --bfile example
```

For technical details of this option, please refer to the paper published in *Am J Hum Genet 83:457-67* [Lou, et al. 2008].

--pii (pedigree-based method II by within family controls)

Option --pii performs the other family-based method, in which phenotypes of siblings are swapped and this method avoids to construct within family pseudo individuals

```
java -jar gmdr.jar --pii --bfile example
```

For the same dataset option --pii runs faster than --pi. For the theory underlying this option, please refer to *Stat Its Interface 4:295-304* [Chen, et al. 2011].

Note: for nuclear families with single affected child, --pii option does not work but --pi does.

--ui (unified method I with family and unrelated samples, exchangeable founders)

For data consisting of family and unrelated samples, to use all of the individuals

```
java -jar gmdr.jar --ui --bfile example
```

in this option, the founders of nuclear families are considered exchangeable to each other as well as to unrelated individuals.

--uui (unified method II with family and unrelated samples, unexchangeable founders)

For data consisting of family and unrelated samples, to use all of the individuals

```
java -jar gmdr.jar --uui --bfile example
```

Differing from option --ui, under option --uui, the founders of nuclear families are only exchangeable to each other as well as to unrelated individuals.

Note: the difference between uui and ui has its realistic implication. For some early developed traits, such as human height, --ui makes more sense than --uui, whereas for late-onset traits or diseases, such as diabetes, --uui option makes more sense.

--x (detect interaction between candidate regions)

--x option enhances flexibility for detecting interactions. We will illustrate its capabilities by examples demonstrated below.

Detecting digenic interaction between SNPs on chromosome 1 and each snp on chromosome 2

```
java -jar gmdr.jar --bfile example --chr 1 2 --x
```

Note: without --x option, this command detect digenic interactions between any two SNPs which are located on chromosome 1 and 2.

Of course, it is straight to detect trigenic interaction on respective SNPs on chromosome 1, 2 and 8

```
java -jar gmdr.jar --bfile example --chr 1 2 8 --x
```

Detecting interaction between a SNP and a chromosome

```
java -jar gmdr.jar --bfile example --snp snp10 --chr 1 2 --x
```

which detects trigenic interaction among snp10 and each SNP on chromosome 1 and each SNP on chromosome 2.

6. Evaluating Empirical p values with permutation

GMDR uses schematic permutation procedures to generate empirically p values. Although computationally intensive, such p values relax assumptions which may not meet in practice. With N rounds of permutation, the empirical distribution will be generated and according to the central limited theory, a z score and its p value can be calculated. The choice of the scheme of permutation is affiliated with the analysis options chosen, and the scheme of the permutation is automated once a

method is chosen.

For the examples used above, the users can entail them with "--perm" option.

```
java -jar gmdr.jar --bfile example --perm 100
```

for case-control samples, evaluating empirical p value for each interaction throughout 100 rounds of permutation.

Once permutation is used, the ordered threshold for testing accuracy will be automatically saved to a file with extension ".thres".

Note: Although more permutation replications reasonably bring out more accurate empirical p values, permutation is very time-consuming. For preliminary analysis, 10 replications of permutation may be a good trade-off of accuracy and cost.

7. SNP selection

Although exhaustive searching of interaction among all SNPs is often desired, it is more realistic to detect interactions among candidate SNPs. GMDR provides a set of options for SNP selection and consequently some computational intensive analysis can be applied to but selected SNPs.

--snp (include or exclude SNPs)

This option provides a couple of ways to include or exclude SNPs in analysis.

Example: detect digenic interactions among snp0, snp1, and snp5

```
java -jar gmdr.jar --bfile example --snp snp0 snp1 snp5 --order 2
```

Example: detect digenic interactions among snp0, snp5

```
java -jar gmdr.jar --bfile example --snp snp0 snp5 --order 2
```

This option may be very useful when the users want to check the interaction between a specific set of SNPs. It greatly saves the time in permutation.

Example: detect interaction between snp0 and snp3-snp5 with 100 permutation

```
java -jar gmdr.jar --bfile example --snp snp0 snp1-snp5 --order 2 --perm 100
```

It will select snp0 and any SNPs within the region, including snp1 and snp5, flanked by snp1 and snp5.

--snp option also can exclude SNPs by adding "--". For the three options

```
java -jar gmdr.jar --bfile example --snp -snp0 -snp1 -snp5 --order 2
```

It detects interaction between the SNPs after excludes snp0, snp1, and snp5.

```
java -jar gmdr.jar --bfile example --snp snp0 snp5 --order 2
```

This option will exclude snps, snp1, snp5 and any snps in between.

```
java -jar gmdr.jar --bfile example --snp -snp1-snp5 --order 2
```

--snpf (include or exclude SNPs through a file)

If the user has a lot of SNPs to include or exclude from the analysis, manually input it at the command line will be very boring. --snpf command reads a file which specifies snp want to include or exclude

For example, if the user has a list of candidate snps from another study, he can import in into gmdr

through --snpf option a file containing

```
“snp0 snp1
snp4-snp7
snp12 snp13”
```

SNPs should be delimited by white space or begins in a new line. Let say the name of the file is snplist.txt, GMDR can import it with the command below

```
java -jar gmdr.jar --bfile example --snpf snplist.txt
```

--chr (include or exclude chromosomes)

```
java -jar gmdr.jar --bfile example --snp snp0 --chr 8 --order 2
```

This option will select snp0 and SNPs on chromosome 8.

```
java -jar gmdr.jar --bfile example --chr 1 8 --order 2
```

Or this command detects digenic interaction for any pair of SNPs on chromosome 1 and 8.

As with --snp option, --chr also can exclude snps simply by adding “-” before the names of chromosomes the user want to exclude

```
java -jar gmdr.jar --bfile example --chr -1 --order 2
```

it excludes chromosome 8 from the analysis.

--window (include SNPs through windows)

```
java -jar gmdr.jar --bfile example --window snp6,5,10 --order 2
```

It includes the SNPs within 5kb at its upstream and 10kb at its downstream of snp5. The distance between SNPs at different chromosomes is considered infinite.

```
java -jar gmdr.jar --bfile example --window snp6,-1,10 --order 2
```

It includes the SNPs located within the upstream of snp6 and located within 10kb downstream. Similarly, if specify the downstream distance to -1, it will include the snps located at the downstream of snp6

```
java -jar gmdr.jar --bfile example --window snp6,5,-1 --order 2
```

--bg (keep certain SNP patten when detecting interactions)

--bg option keeps certain interaction patterns into a model as background.

```
java -jar gmdr.jar --bfile example --bg snp0 snp1 --order 2
```

It detects digenic interactions of snp0 and snp1 to the rest of SNPs.

```
java -jar gmdr.jar --bfile example --bg snp0 snp1 --chr 8 --order 3
```

It detects trigenic interactions of snp0 and snp1 with any single SNPs on chromosome 8.

Note: --bg applies for SNPs only.

--maf (threshold for excluding SNPs)

This option specifies threshold for minor allele frequency in excluding SNPs. For example, to exclude SNPs of which MAF are less than 0.1

```
java -jar gmdr.jar --bfile example --maf 0.1
```

Notes: allele frequencies are calculated based on selected individuals in the analysis rather than always the whole data. For different methods, such as “cc”, “pi”, “pii”, “ui”, and “uii”, specified, the allele frequencies may differ.

8. Individual selection

--exfam (exclude families)

Excluding families from the analysis, for example, to exclude families with id 1 and 2 in the example file

```
java -jar gmdr.jar --bfile example --exfam 1 2
```

--exfamf (exclude families through a file)

Similarly, the user can put all families want to excluded into a file.

--exnosex (exclude individuals whose sexes were unknown)

This option excludes individuals whose sexes are unknown or neither coded as 1 (male) nor as 2 (female).

--male (include males only)

This option includes males (coded as 1) only.

--female (include females only)

This option includes females (coded as 2) only.

--geno (threshold for genotype missing rates)

Excluding certain individuals by specifying missing genotype rate

```
java -jar gmdr.jar --bfile example --geno 0.1
```

any individual whose genotype missing rate of the selected markers is higher than 0.1 will be excluded from the analysis.

Notes: as with option maf, genotype frequencies are calculated based on selected individuals in the analysis rather than always the whole data. For different methods, such as “cc”, “pi”, “pii”, “ui”, and “uii”, specified, the genotype frequencies may differ.

9. Phenotypes and covariates

If the user wants to analyze other phenotypes rather than affection status as listed in the sixth column in PED/BED file, a phenotype file should then be loaded. In this section, we will introduce options related to the operation of phenotypes and covariates.

--phenotype (specify a phenotype file)

Load your phenotype file through this option, the first line of a phenotype file is the title line and the

first two columns are mandatory:

Family ID

Individual ID

Without loss of generality, a phenotype file often looks like below

```
FID ID phe0 phe1 phe2
1 10000 0 0.8875 0.1125
1 10001 0 0.875 0.125
1 10002 0 0.87 0.13
1 10003 1 0.88 0.12
```

If want to load example.phe and choose phe0 as the response and phe1 as the covariate, the command can be

```
java -jar gmdr.jar --bfile example --pheno example.phe --responsename phe0 --covarname phe1
```

or use the indexes of the columns (note that the first two columns are accounted, the third column in fact is considered the first column), bus should use --response and --covar options instead.

```
java -jar gmdr.jar --bfile example --pheno example.phe --response 1 --covar 2
```

--reg (method for building score statistics)

By default, linear regression is used to build score

```
java -jar gmdr.jar --bfile example --pheno example.phe --response 1 --covar 2 --reg 1
```

Or, if the user still want to use affection status listed in the PED/BED file as the response 0

```
java -jar gmdr.jar --bfile example --pheno example.phe --response 0 --covar 2 --reg 1
```

or simply do not bother --response and use --covar to specify covariates.

```
java -jar gmdr.jar --bfile example --pheno example.phe --covar 2 --reg 1
```

for more than 1 covariate, the user can specify them one by one and delimit them with white space,

```
java -jar gmdr.jar --bfile example --pheno example.phe --covar 2 3 --reg 1
```

or for a range of covariates by using "--"

```
java -jar gmdr.jar --bfile example --pheno example.phe --covar 2-3 --reg 1
```

Notes: either --response or --responsename can take one parameter only.

10. Cross-validation options

--cv (fold for cross-validation)

Specifies the folder for cross-validation, by default, cv is set 5. The user can set it to other integer but should be not less than 2. For example, to set cv to 10, use the command listed below

```
java -jar gmdr.jar --file example --cv 10
```

Note: a bigger value of cv corresponds to more computing time.

--seed

A seed value is required in GMDR for stochastic procedures involved, and by default, seed is set 2011. The analysis results may differ when a different seed is used.

--tie (classification for a tie group)

Specifies the classification for a genotypic combination in tie. By default, GMDR classifies a tie genotypic cell as high-risk (--tie h), but the user can switch it to low-risk by specifying "--tie l", or exclude such any genotypic combinations in tie by specifying "--tie -1".

11. Output options

--out (the root name for the output files)

By default, the result will be saved into such as "gmdr.txt" for interaction, and "gmdr.thres" for permutation. The user can specify the preferred root name of the output files

```
java -jar gmdr.jar --file example --out mygmdr
```

Then the output names will be "mygmdr.log", "mygmdr.int", and "mygmdr.thres", respectively.

--verbose (complete the output information)

To view the complete information of an interaction this option can be turned on. Then the output information will include minor allele frequencies, classification of each genotypic combination.

```
java -jar gmdr.jar --file example --verbose
```

When in brief format, the output information includes:

- model
- effective individuals,
- vc(vx, vt);
- Testing Accuracy,
- Training Accuracy;

Notes: For testing accuracy, once permutation is used, its empirical mean, variance and p value will be listed. vx means the variance explained, vt means the variance of the trait, and $vc = vx / vt$.

When in verbose format, the output information includes:

- model
- effective individuals
- vc(vx, vt)
- Testing Accuracy
- Training Accuracy
- locus1, chr1, position1, MAF1
- classification (genotype, risk group, positive scores, positive subjects, negative score, negative subjects).

Notes: for risk group, H for high-risk and L for low-risk.

--train (threshold for training accuracy for output)

Specify the threshold for training accuracy, only interactions with training accuracy greater than the threshold will be printed out. Empirically, training accuracy is often around 0.5~0.54, and a strong

interaction often has training accuracy above 0.54 or greater.

--test (threshold for testing accuracy for output)

Specify the threshold for testing accuracy, only interactions with testing accuracy greater than this threshold will be printed out. Empirically, training accuracy is often around 0.45~0.53, and a strong interaction often has testing accuracy above 0.53 or greater.

--p (threshold for p value for output)

Specify the threshold for empirical p values evaluated by permutation. This option only works when --perm option is specified

--vc (threshold for variance contributed for output)

Specify the threshold for variance contributed by interactions. Variance contributed is measured as the percent of the variance explained compared with the total variance of a trait. For a trait collected without selection, variance contributed is heritability, whereas it is not if the trait is collected through specific schemes. It should be noted that for complex traits contributed variance is often very small (1%~5%).

All these four thresholds, for --p option permutation should be in effect, can be used together. For example

```
java -jar gmdr.jar --file example --train 0.53 --test 0.53 --p 0.05 --vc 0.02 --verbose --perm 30
```

12. Estimate the time required to complete the analysis

--testdrive

It is well-known that detecting gene-gene interactions could be very time-consuming, especially toward high-order interaction with large datasets. To estimate the time required to complete the whole analysis will help user make a more realistic analysis strategy before load a huge project into GMDR. GMDR introduce --testdrive option to assist the user estimating the computation load of the project. Once a the testdrive option is specified, GMDR will run the first 1000 interactions and based on the time used to give an estimated time for the whole analysis.

```
java -jar gmdr.jar --file example --perm 100 --order 4 --testdrive
```

It reports the estimated time for completing the analysis project.

Notes: when using testdrive option, GMDR quits right after finishes the analysis on the first 1000 interactions. The time for loading data, which can be substantial, is excluded from the estimated time.

13. Reduce computational burden

--thin

When the user only wants to randomly check the interaction rather than exhaustively, --thin option can

be very useful. “--thin 0.2” samples 20% of in total interactions,

```
java -jar gmdr.jar --file example --thin 0.2
```

it can dramatically reduce computation burden. Or, the user can further applies permutation to this command

```
java -jar gmdr.jar --file example --thin 0.2 --perm 100
```

14.GMDR and HPC

Once the user is going to run a big GMDR project, which may be very time-consuming, it is suggested to implement the analysis at a cluster. In this way, the user can first partition the analysis into slices and analyze each independently to a computing node.

--slice

```
java -jar gmdr.jar --file example --perm 100 --slice 1/5
```

As now the whole analysis has been partitioned into five slices, the output files will entail this information and read “gmdr1.txt.slice1.5” and “gmdr1.thres.slice1.5”.

For example, assuming 1 million interaction patterns to test, the user can partition it into, say 5 slices, and run them on the cluster.

```
java -jar gmdr.jar --file example --slice 1/5
```

```
java -jar gmdr.jar --file example --slice 2/5
```

```
java -jar gmdr.jar --file example --slice 3/5
```

```
java -jar gmdr.jar --file example --slice 4/5
```

```
java -jar gmdr.jar --file example --slice 5/5
```

From the first to the fifth commands, the whole interaction analysis is partitioned into 5 slices and each slice is implemented independently once these scripts are submitted to a cluster. The simplest way is to include each of these five commands into an independent job script, then it can be run in five nodes on a cluster. It will reduce the time by 80%. Typically, a pbs script reads like this:

```
#!/bin/bash
```

```
#$ -S /bin/bash
```

```
#$ -cwd
```

```
#$ -N GMDR_Slice1
```

```
#$ -l h_rt=10:10:00,s_rt=10:08:00,vf=1G
```

```
#$ -m eas
```

```
# Tell the scheduler to use the environment from your current shell
```

```
#
```

```
#$ -V
```

```
java -jar gmdr.jar --bfile example --slice 1/5
```

15. Miscellanea

--ss

If the affection status listed in the sixth column of the PED/BED is 2 for affected and 1 for unaffected, --ss option should be specified. Otherwise gmdr may mistake the affection status.

--missingallele

Specify the code for missing alleles. By default, it takes value of "0".

--missingpheno

Specify the code for missing values in the phenotype file. By default, it takes value of "-9".

16. References

- Chen GB, Zhu J, Lou XY. 2011. A faster pedigree-based generalized multifactor dimensionality reduction method for detecting gene-gene interactions. *Stat Its Interface* 4(3):295-304.
- Lou XY, Chen GB, Yan L, Ma JZ, Mangold JE, Zhu J, Elston RC, Li MD. 2008. A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. *Am J Hum Genet* 83(4):457-67.
- Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, Li MD. 2007. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet* 80(6):1125-37.

17.List of options

Option	Parameter/default	Description
File		
--file		Specify .ped and .map files if they have the same root name.
--ped		Specify .ped file.
--map		Specify .map file.
--bfile		Specify .bed, .bim and .fam files.
--bed		Specify .bed file.
--bim		Specify .bim file.
--fam		Specify .fam file.
Phenotype and covariates		
--pheno		Specify phenotype file.
--response	-1	Specify response by number
--responsename	null	Specify response by name
--covar	null	Specify covariate by number
--covarname	null	Specify covariate by name
--reg	0	Specify either linear or logistic regression for calculating score statistics. 0 for linear regression, 1 for logistic regression.
Detection of Interactions		
--perm		Specify the number for permutation
--cc		"cc" for case-control design, which extract unrelated individuals only.
--ui		"ui" for the unified model, in which the information of the founders and the children will be combined altogether. Founders are considered exchangeable to each other
--uii		"uii" for the unified model, in which the information of the founders and the children will be combined altogether. Founders are considered exchangeable to each other only within a family.
--pi		"pi" implements the model that is conditional on parental genotypes as proposed in AJHG2008.
--pii		"pii" implements the model that is conditional on

		parental genotypes proposed in <i>Statistics and Its Interface</i> .
--x		Interaction for specified snp lists. Mark: when this option is turned on, --order option is unable. The dimension of the interaction searched for is determined by the group of selected SNPs. For example, --snp snp1 snp2 snp3-snp7, will detect interaction of snp1, snp2 and any snp in the group defined by snp3-snp7. If want to detect digenic interaction between the regions 10kb around snp10 and snp100 snps around snp10 and snp100, use "--snpwindow snp10,10,10 snp100,10,10 --x". If want to detect interaction between snps on chromosome 1 and 2, use "--chr 1 2 --x". If want to detect interaction of a single snp to a chromosome, use "--chr 1 --snp snp1 -x" or "--chr 1 -bg snp1 -x".
SNP selection		
--window		Specify snp and its window. snp1,10,10, the 10kb region flanking snp1 snp2,10,-1, the 10kb ahead of snp until the end of the chromosome in which the snp2 is. snp3,-1,10, selecting all SNPs located on the same chromosome where snp3 is until the 10kb down away from it.
--snp		Specify snps included or excluded, "--snp snp1 snp3-snp5 snp7 -snp8". If a snp is both included and excluded, it will be excluded.
--chr		Specify chromosomes included or excluded, "--chr 1 -2" includes chromosome 1 while excludes 2.
		Mark: --snpwindow, --snp, and chr can be combined together. If each snp only included even it is selected more than once. After selection, only the snps selected will be read into the memory. If the regions for analysis is clear, it is suggested to use this option.
--bg		Specify the background snps. Only single snps are allowed.
		Mark: snps specified in --bgsnp will always be included when detecting gene-gene interactions.

		If want to look interaction between snp1 and snp2, use option --bgsnp snp1 snp2.
--maf		Specify MAF threshold below which will be excluded
Individual selection		
--exfam		Specify excluded families : --exfam 1 2
--exfamf		Specify the file containing excluded families: one family per line.
--geno		Specify the maximum per-person genotype missing
		Mark: Neither maf nor geno will check the snps included in --bgsnp. When --x is turned on, neither maf nor geno will check the snps included.
MDR options		
--cv	5	Specify cross-validation
--seed	2011	Specify the seed for stochastic processes.
--order	1	Specify the order of interaction searched for
--tie	Default as high risk for a tie genotype	Specify the classification for tie genotypic cells. As default setting, "-- tie h" classifies a tie genotypic cell high risk; "--tie l" classifies a tie genotypic cell low risk; otherwise, classify it as unknown.
--thin		Specify the random sampling fraction between 0~1. 0.2 means 20% of interactions will be sampled for detecting gene-gene interaction.
--slice		Specify the partition of the searching space. --slice 2/10, it searches the second slice of the searching space which is partitioned into 10 slices. Mark: this one can be used to partition large project into slices and load each one into cluster.
		Mark: both "thin" and "slice" options help to reduce computing load, but there are some differences. "Thin" reduces the loading by sampling that eventually gives up exhaustive search strategy. But the slice option can iterate all interaction spaces, but only calculate the specified part. If the searching space is extremely huge, high performance computing with "slice" can

		be very helpful. If there are 100 CPUs available, the user can partition the interaction searching space into 100 slices, and load each slice into an independent node, “—slice 1/100” for the node 1, “—slice 2/100” for the node 2, etc. It can significantly boost the whole analysis.
Output		
--missingallele	Default “0”	Specify the code for missing alleles.
--missingphenotype	Default “-9”	Specify the code for missing phenotypes.
--ss		Specify the shift for status if affected and unaffected are coded 2 and 1.
		Mark: individuals with missing value for the selected phenotype or covars will be eliminated from the analysis.
Output		
--verbose		Print the complete results for each interaction.
--train		Specify the threshold of training accuracy for output.
--test		Specify the threshold of testing accuracy for output.
--p		Specify the threshold of empirical p values for output. This option works only when perm option is specified.
--vc		Specify the threshold of variance contributed for output.